Definitions and Measures of Intelligence in Deep Blue and The Army XUV

John M. Evans, Ph.D.
John M Evans LLC
1 Reservoir Road
Newtown, CT 06470
john_evans@snet.net

Abstract**—**Three different definitions of intelligence are reviewed, using Deep Blue as the basis for comparison, and a discussion of chess rating points as a metric of performance is presented. It is argued that Deep Blue showed intelligent behavior and passed a form of the Turing Test. Further applications of similar search techniques in the Army XUV are shown to generate behaviors that show elements of intelligence in autonomous systems.

Keywords**:** *definitions of intelligence, measures of intelligence, computer chess, Deep Blue, autonomous vehicles, path planning, Army XUV*

I. Deep Blue

A remarkable milestone in computer science was achieved in 1997 when IBM's computer, Deep Blue, beat the World Champion, Garry Kasparov, at chess [1,2]. This was the culmination of fifty years of work on what was considered a problem that "penetrated to the core of human intellectual endeavor."[3] This event ranks with the Wright Brothers first flight and with the achievement of sustained fission in the Manhattan Project: early success with what became (or in this case will become) world-changing technologies.

It is interesting to note that Hsu, the designer of the VLSI chips used in Deep Blue, characterizes the matches with Kasparov not as man versus machine but rather as man as performer (Kasparov) versus man as toolmaker (Hsu and the IBM team). [1] Deep Blue was a remarkable tool, quite successful at the specialized task for which it was designed.

II. Definitions of Intelligence

Was Deep Blue "intelligent"? Artificial Intelligence researchers generally say no, that Deep Blue's success depended on special purpose chips that were designed only for evaluating chess moves. Philosophers and psychologists generally say no, that there was no self-awareness, no consciousness, no real understanding in Deep Blue. However, the English mathematician Alan Turing in 1950 proposed an operational definition of "intelligence" that basically said that if a person interacting with another unknown entity could not distinguish between a computer and a person, then that entity would have to be considered intelligent. [4,5]

The Turing Test defines "intelligence" in terms of black box functionality of a machine in comparison with a human in human/machine interaction. Searle, with his famous Chinese Room argument, redefines intelligence in terms of understanding. [6, 7] Hawkins, in his book On Intelligence, defines intelligence in terms of predictive ability. [8] These three definitions are not necessarily incompatible, although Searle was specifically attacking the Turing Test definition.

Deep Blue had all of these facets of intelligence. Kasparov felt he was not playing with a machine but with an independent intelligence, an entity with an independent mind: "Now for the first time we are playing not with a computer, but with something that has its own intelligence."[9] He made this statement after winning game 2 of the first match in Philadelphia in 1996. In a less charitable mood during the losing rematch in New York in 1997, he accused IBM of cheating, of having a person directing the game. [10] From Kasparov's standpoint, Deep Blue passed a version of the Turing Test.

III. The Structure of Deep Blue

Deep Blue also had predictive capability (Hawkin's definition of intelligence) and embodied understanding of the game of chess (Searle's definition). To understand this we must delve into how Deep Blue operated.

The basic principles of playing chess with a computer were laid out by Turing in England and Shannon in the U.S. by 1950 [11, 12]. Since then there has been refinement and the addition of bells and whistles, but the main point is that increasing computer power makes it possible to look farther ahead in a game and that in turn leads to greater skill and therefore greater perceived "intelligence".

The chess game for a computer is divided into three parts. [1,2] The first part is the opening book, a sequence of scripted moves that have been played out many times in the past. This is essentially table lookup.

The second part of the game uses search techniques to evaluate different possible moves. At each level of search a quantitative value is calculated based on material and board positions and the search is selectively deepened along the most promising lines. The weights given to different pieces and different board positions were developed with the help of grandmasters and it is in these evaluation functions that chess knowledge is embedded in Deep Blue. Deep Blue calculated an average of fourteen plies (half-moves) but in some cases went to twenty-ply or even thirty-ply deep evaluations in examining possible lines of play. [1,2]

The final part of the game is the endgame. Deep Blue had all four piece endgames stored in main memory and all five piece and some six piece endgames stored on disk. At this point it is again a table lookup strategy.

In the first and last part of the game, understanding of the game of chess is embedded in the "book", the tables, which were created by human grandmasters. In the middle part of the game, understanding of chess is embedded in the evaluation functions, which again were set by grandmasters working with the chip designers and the programmers. In terms of Searle's argument, Deep Blue did not understand what was going on when it executed the evaluation functions, but it did ***embody*** understanding of the game of chess, and hence, from Kasparov's viewpoint, it seemed to possess intelligence.

From the standpoint of Hawkins definition, Deep Blue was exhibiting intelligent behavior by being able to predict the future results of its actions. This is the essence of cost-based search.

Cost based search is not how humans play chess at the highest levels. [13] Instead, we exploit the massive parallel processing capability of the human mind together with the basic ability of the neocortex to recognize and store patterns and sequences of patterns (8) and play mostly on the basis of pattern recognition. Functional MRI tests show that chess experts activate primarily the parietal cortex, where spatial patterns are stored, while novices activate primarily the temporal cortex, where information on individual pieces and their capabilities are stored. Either approach, pattern recognition with massively parallel processing or search with a Von Neumann architecture computer, obviously works.

IV. Performance Metric for Chess

In the case of chess there is an established performance metric, Elo's chess rating points system [14], adopted by FIDE (Fédération Internationale des Échecs) as the official international rating system for chess players. This is a statistical rating based on head-to-head competitions: a player taking 3 out of 4 points in a four game match is considered 200 rating points above the losing player. Newborn has developed experimental data based on running real chess matches through computer chess programs. His data indicate that increasing the depth of search by one level gives an increase of at least 100 rating points in skill level. [15] It is then just a matter of applying sufficient computing power to reach a level of skill beyond that of any human. Deep Blue was able to exercise the equivalent of two to three trillion instructions per second (using hundreds of special purpose VLSI chips on a cluster computer with 36 processors) which allowed it to examine two hundred million board positions per second and it played in New York at a rating level of over 2800, on a par with Kasparov [1,2].

The following data on ratings for specific chess playing computers and computer programs is from Newborn [2].

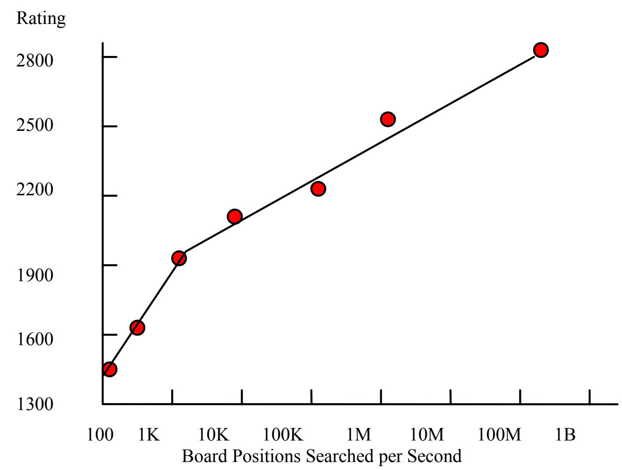

Figure 1: Rating vs Speed of Search

V. The Army XUV

Another interesting example of using cost-based search to generate complex behaviors is the vehicle control for the Army's Experimental Unmanned Ground Vehicle Program, commonly referred to as Demo III, which ended in 2003. [16, 24] (Demo I was teleoperation, Demo II was supervised autonomy, and Demo III was targeted at full autonomy for scout missions of reconnaissance, surveillance, and target acquisition). Many of the ideas from Demo III were embodied in the Stanford and Carnegie Mellon winners of the DARPA Grand Challenge Road Race in 2005 and this technology is now being developed into the Autonomous Navigations System for the Army's Future Combat Systems vehicles.

This special vehicle, shown in Figure 2, has four wheel hydraulic drive with four wheel steering. The navigation sensors include scanning ladar, stereo cameras and stereo FLIRs, microwave radar, bumpers, tilt sensors, GPS and inertial navigation. The mission sensors are in the dome package on a shock mount on top of the vehicle.

The path planning for Demo III used cost based search. [17, 18, 19, 21] The cost function is

$$\text{Cost} = \mathbf{\Sigma}\, c_i * v_i$$

Figure 2: Army XUV (Experimental Unmanned Ground Vehicle)

where the c_i are relative costs or weights (relative importance) of different relevant state variables and v_i are the current values of those variables. There were 16 variables used in the cost function, including side slope, forward to back slope, ground roughness, ground center height, soil properties, on-road, off-road, vegetation, obstacles, and mission completion time.

Cost maps were created using a priori knowledge (maps) and real time terrain knowledge from vision and ladar scanners. A grid of points was cast onto the maps and points were connected to provide possible path segments. A search was then conducted, calculating the cost for each path segment encountered and deepening the search along favorable directions to find the lowest cost path from a starting point to a finishing point. Search was carried out at several levels of resolution: 5 m, 50 m and 500m range maps. The 500 m maps gave optimal start and end points for the 50 m maps, which in turn gave start and end points to the 5 m map. At the 5 m map level, pre-calculated trajectories that embodied vehicle dynamics (speed, inertia, possible steering rates, etc.) were used for computational efficiency instead of an arbitrary point grid. [19]

Neural nets, which model how our brains work, carry out exactly this type of computation, summing the product of neural signal strengths times synapse weights. As Churchland and Sejnoski note this neatly wraps vector representation with compatible matrix processing and allows for many types of mental computations. [20] We use chemistry to create emotions to modify weights for low level behaviors (e.g. fight or flee) and logical computation for more abstract behavior generation.

The following figures, developed by Balakirsky [21] show examples of behavior generation for the Demo III vehicle under different weight assumptions. The figures are from a simulated operator control unit and use data from topological maps of the grounds of the National Institute of Standards and Technology (NIST). The NIST maps were only course resolution, so the images seem blurred. Red areas are obstacles (trees, bushes and fence lines), blue are buildings, and green are roads and parking lots.

Figure 3 shows a hypothetical path from one point to another with the weight for obstacles being high and all other weights being low. The faint white line is a direct line from the start (lower right) to the finish (just beyond the road in the upper left) and the heavier yellow line shows the computed path. The vehicle finds an opening through the trees and then returns to the direct path to the goal.

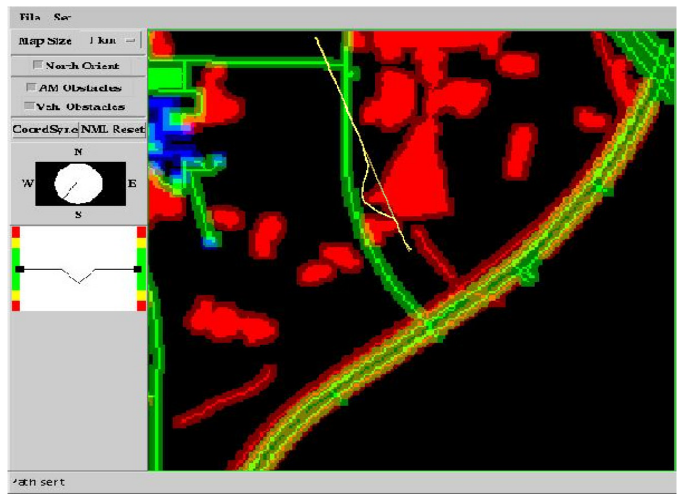

Figure 3: Simulated path planning with obstacles having high weights

In Figure 4 the weights for being off-road are set very high, so the vehicle does not go directly toward the goal but instead heads toward the nearest road and then stays on roadways until it is as close as it can get to the goal, at which point it departs from the road and dashes to the finish.

Finally, in Figure 5, the weight for being out in the open and hence potentially detectible by an enemy force is set very high and the cost for mission completion time is set low. The result is stealth behavior, running carefully along the tree line to stay under overhanging branches as much as possible. This is a tactically significant behavior for an Army scout, and it was generated by the wonderfully general approach of cost based search that matches how our own minds determine appropriate behavior.

Future work will include learning and recalling appropriate weights for different situations. The ability to recognize contexts and select appropriate weights depends in turn on improved perception to generate image understanding and situational awareness. The Demo III program only scratched the surface of these issues.

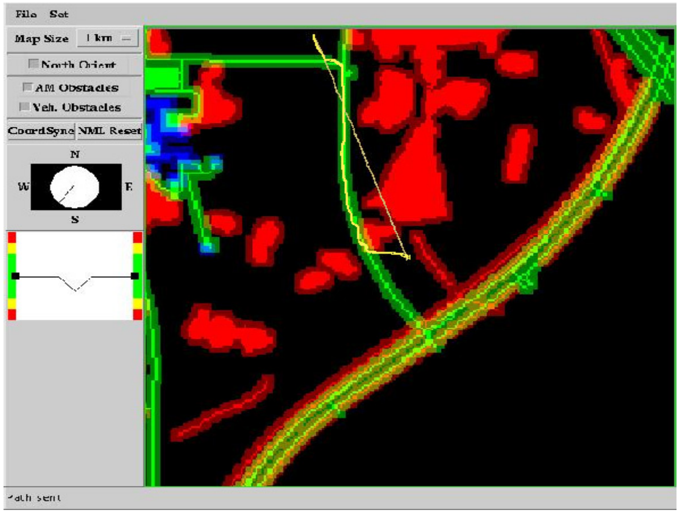

Figure 4: Path Planning with low cost for On-Road and high cost for Off-Road

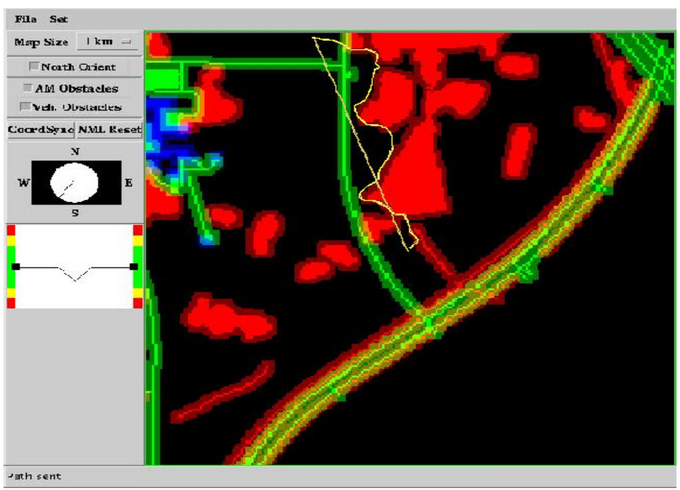

Figure 5: High Cost for being in the open (detectable by an enemy scout) and Low Cost for mission completion time.

VI. Performance Metrics

There is not single metric for performance of unmanned vehicle planning systems to match the chess rating system. Instead the focus has been on measuring the performance of unmanned systems in navigating known courses of various difficulties. Measures of performance include completion of segments of the course, time to completion, number of targets found in a course and number of interventions. [22, 23, 24] For smaller robots artificial test courses have been used for evaluation in both simulation and in contests. [23, 25].

VII. Conclusion

This paper has examined some interesting idiot savant capabilities: Deep Blue beat Kasparov and must be considered intelligent in its domain, but all it does is play chess. The Army XUV showed some intriguing elements of intelligent behavior, but it was badly nearsighted, had very limited perceptual understanding and was computationally bound, so it also showed some truly dumb behaviors at times and got lost on a regular basis. Still, the techniques of planning and problem solving, one aspect of intelligent robot systems, are now fairly well understood, objective test methods and metrics are being developed to benchmark capabilities and rapid progress is being made.

References

[1] Hsu, Feng-Hsiung, *Behind Deep Blue*. Princeton: Princeton University Press, 2002.
[2] Newborn, M., *Deep Blue*. New York: Springer-Verlag, 2003.
[3] Newell, A., Shaw, C. and Simon, H., from an early paper on chess, as quoted in Hsu, *op.cit.*
[4] Turing, Alan, “Computing Machinery and Intelligence”. *Mind*, **LIX**, pp433-460, 1950.
[5] Hodges, A., *Turing*. New York: Routledge, 1999.
[6] Searle, J., *Minds, Brains and Science*. Cambridge: Harvard University Press, 1984.
[7] Searle, J., “Is the Brain’s Mind a Computer Program?” *Scientific American*, **262** (1), 1990.
[8] Hawkins, J., *On Intelligence*. New York: Henry Holt & Co., 2004
[9] Kasparov, Garry, as quoted in Newborn, *op. cit.*
[10] Kasparov, Garry, as quoted in Hsu, *op. cit.*
[11] Turing, Alan, “Digital Computers Applied to Games”, in *Faster Than Thought*, B.V. Bowden (ed.), London: Pitman, 1953.
[12] Shannon, Claude,”Programming a Computer for Playing Chess”, *Philosophical Magazine*, **41**(7), 1950.
[13] Ross, Phillip E., “The Expert Mind”. *Scientific American*, August, 2006.
[14] Elo, A., conceived of the statistical performance rating system used in chess. http://en.wikipedia.org/wiki/ELO_rating_system.
[15] Newborn, *op. cit.,* Appendix H.
[16] Shoemaker, C., et al., “Demo III: Department of Defense Testbed for unmanned Ground Mobility”, *Proc. SPIE Conference on Unmanned Ground Vehicle Technology*, SPIE Vol. 3693, Orlando, FL, April, 1999.
[17] Albus, J., et al., *4D/RCS Version 2.0: A Reference Model Architecture for Unmanned Vehicle Systems*, NISTIR 6910, Gaithersburg, MD, 2002.
[18] Schlenoff, Madhavan, R., and Balakirsky, S., “Representing Dynamic Environments for Autonomous Navigation”, *Proc. IEEE Int. Conf. on Intelligent Robots and Systems*, 2003.
[19] Balakirsky, S. and Lacaze, A., “World Modeling and Behavior Generation for Autonomous Ground Vehicle”, *Proc. IEEE International Conference on Robots and Automation*, 2000.
[20] Churchland, P. and Sejnoski, T., *The Computational Brain.* Cambridge: MIT Press, 1999.
[21] Balakirsky, S., *A Framework for Planning with Incrementally Created Graphs in Attributed Problem Spaces*, Berlin, Germany: IOS Press, 2003.
[22] Messina, E. and Meystel, A., *Proc. PerMIS* 2000-2004. Gaithersburg, NIST SP 982, 990, 1014, 1036, 2000-2004.
[23] Jacoff, A., Messina, E., and Evans, J., “Performance Evaluation of Autonomous Mobile Robots”, *Industrial Robot*, **29**:3, May 2002.
[24] Madhaven, R., Messina, E., and Albus, J., *Intelligent Vehicle Systems*. New York: Nova Science Publishers, 2006. Chapter 3 covers behavior generation; chapter 9 covers performance evaluation; chapter 10 is a review of the Army Demo I-Demo III programs.
[25] Jacoff, A., Weiss, B., and Messina, E., “Evoluation of Metrics and Performance of USAR Competitions”, *Proc. 2003 PerMIS*, Gaithersburg, MD: NIST SP 1014, 2003.